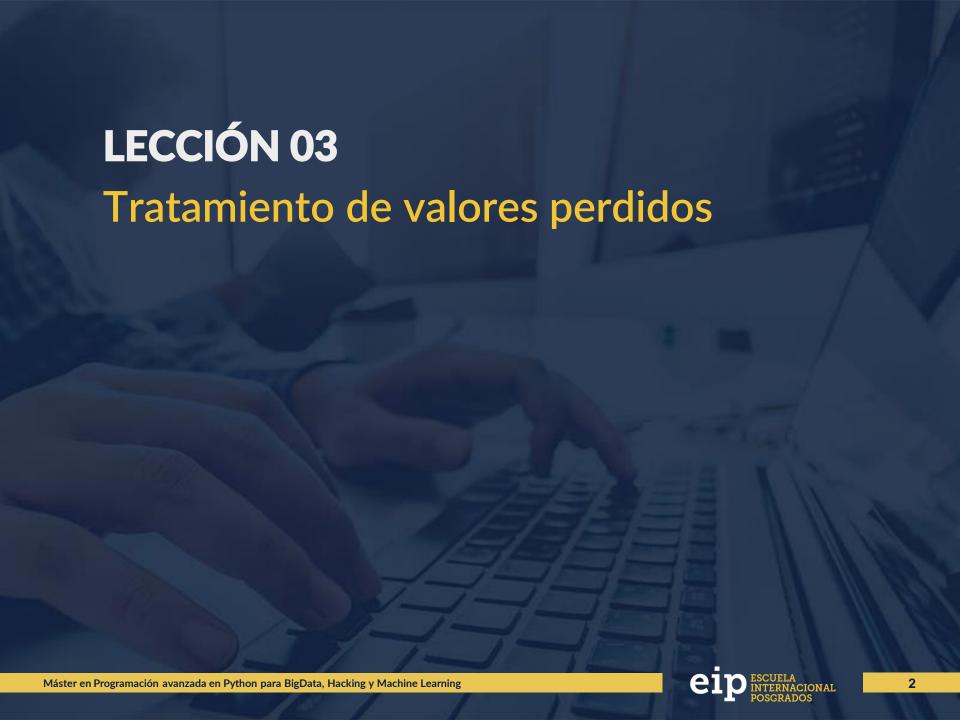


Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Programación Python para Machine Learning



ÍNDICE

- ✓ Introducción
- Objetivos
- ✓ Valores perdidos
- Estrategias de tratamiento de valores perdidos
- Conclusiones

OBJETIVOS

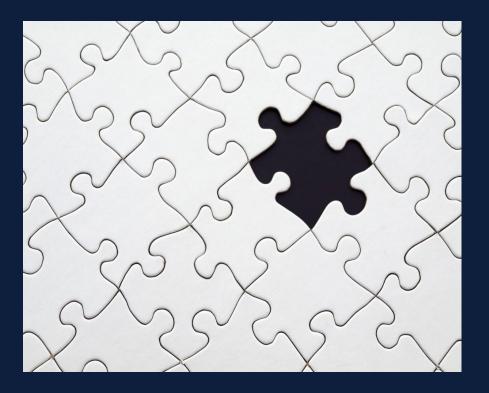
Al finalizar esta lección serás capaz de:

- 1 Entender la problemática de la presencia de valores perdidos.
- 2 Identificar de modo adecuado los valores perdidos en un conjunto de datos.
- Aprender a utilizar las técnicas de eliminación de valores perdidos.
- 4 Manejar las técnicas principales para la imputación de valores perdidos.



INTRODUCCIÓN

- ✓ Los procedentes de problemas en el mundo real no son perfectos.
- ✓ Valores faltantes.
- ✓ Valores atípicos.
- ✓ Conocimiento del problema para su tratamiento.



MARCAR VALORES PERDIDOS

Máster en Programación avanzada en Python para BigData, Hacking y Machine Learning

Detección de valores perdidos.

Filosofía Pandas

	BI-RADS	Age	Shape	Margin	Density	Severity
0	5	67	3	5	3	1
1	4	43	1	1	NaN	1
2	5	58	4	5	3	1
3	4	28	1	1	3	0
4	5	74	1	5	NaN	1
	• • •		• • •			

ESTRATEGIAS PARA TRATAR CON LOS VALORES PERDIDOS

La presencia de valores perdidos en un conjunto de datos es habitual.

Hay que abordar su tratamiento.

Estrategias:

✓ Eliminar

✓ Imputar

	BI-RADS	Age	Shape	Margin	Density	Severity
0	5	67	3	5	3	1
1	4	43	1	1	NaN	1
2	5	58	4	5	3	1
3	4	28	1	1	3	0
4	5	74	1	5	NaN	1
	• • •		• • •	• • •	• • •	

Eliminar instancias o características con datos perdidos

Estrategia más sencilla, pero tiene sus inconvenientes.

Distintas posibilidades:

- Eliminar filas con al menos un valor faltante.
- Eliminar las filas en las que todos los valores son faltantes.
- Eliminar las columnas con al menos un valor faltante.
- Eliminar las columnas con todos los valores faltantes.
- Eliminar filas/columnas cuando al menos haya 'n' valores faltantes.

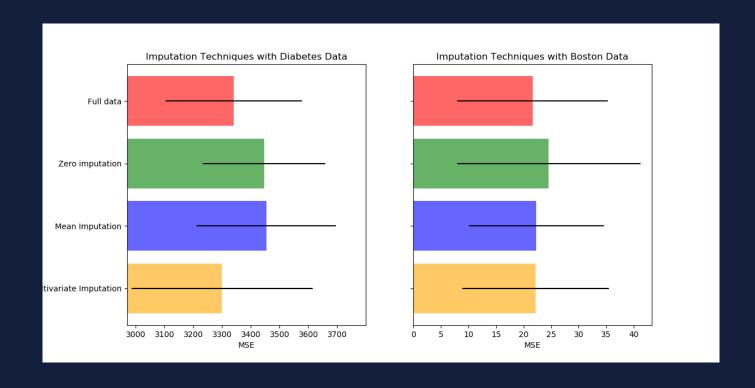
Imputación de valores perdidos

Imputación: uso de una técnica para reemplazar los valores perdidos.

Distintas posibilidades:

- Imputar un valor constante que tenga significado dentro del dominio.
- > Imputar un valor de otro registro seleccionado.
- Imputar un estadístico representativo de la columna.
- Imputar un valor estimado por otro modelo predictivo.

Comparativa de técnicas



MUCHAS GRACIAS POR SU ATENCIÓN











